A STUDY ON SOUND SOURCE APPARENT SHAPE AND WIDENESS

Guillaume Potard, Ian Burnett

School of Electrical, Computer and Telecommunications Engineering
University of Wollongong
Northfields Avenue
Wollongong, NSW 2500, Australia
{gp03,ian_burnett}@uow.edu.au

ABSTRACT

This work is intended as an initial investigation into the perception of wideness and shape of sound sources. A method that employs multiple uncorrelated point sources is used in order to form "sound shapes". Several experiments were carried out in which, after some initial training, subjects were asked to indentify the shapes that were being played. Results indicate that differences in vertical and horizontal source wideness are easily perceived and scenes that use broad sound sources to represent normally large sound objects are selected 70% of the time over point source versions. However, shape identification was found to be more ambiguous except for certain types of signals where results were above statistical probability. The work indicates that shape and wideness of sound sources could be effectively used as extra cues in virtual auditory displays and generally improve the realism of virtual 3D sound scenes. This work was performed as a Core Experiment within the MPEG Audio Subgroup with the intention of possible integration of source wideness into MPEG-4 AudioBIFS.

1. INTRODUCTION

The apparent width or extent of natural sound sources is an important perceptual cue which is, however, rarely implemented or properly controlled in virtual 3D sound scenes. Natural sound sources such as a beachfront or a waterfall usually exhibit a particular spatial extent which provides information on the physical dimensions, geometry and distance of the sound emitting object. Accurate control of the spatial distribution of a sound source can thus be an efficient way to improve the realism of 3D sound scenes as well as providing extra dimensions in the process of data sonification.

AudioBIFS [1], the virtual sound scene description scheme of MPEG-4, can currently only define point sound sources such as a flying insect or a distant sound source. The resulting lack of spatial extent in sound sources can, in some cases, severely limit the realism of the virtual sound scenes. We therefore carried out psychoacoustic experiments aimed at deriving the necessary parameters to express wideness and shape of sound sources in virtual sound scenes.

The first experiments were aimed at determining if listeners could discriminate between particular sound shapes. The sound shapes were created using multiple point sound sources emitting uncorrelated signals.

In the initial experiments, real sound sources (i.e. speakers) were used for the point sources. However, in later work, virtual sound sources were employed using Ambisonics [2] spatialisation technique on a cubic speaker array. The experiments were repeated for several types of noise and a blues guitar sample.

One dimensional vertical and horizontal sound source wideness were also studied with final experiments determining if virtual sound scenes using broad sound sources to represent large auditory objects or events (e.g. beach, thunder etc.) were more natural than those based solely on point sound sources.

Before we consider the detailed experimental results, the following section gives a little background on the areas of broad and shaped sound sources.

2. BACKGROUND

Research on sound source wideness, also known as tonal volume, began in the 20s [3] and found that the perceived wideness of a single sound source is a function of signal frequency, loudness [4] and signal duration [5].

Low pitched sound sources need a greater distance for one wavelength to unfold and tend to have a larger apparent width than high pitched sound sources.

Loudness, which is typically inversely proportional to the distance between the sound source and the listener, also influences perceived wideness. In effect the apparent width of a sound source decreases with distance in the same way that visual objects appear smaller when distant.

Apparent source width is also tightly linked to the Inter Aural Cross Correlation (IACC) value [6][7]. In concert halls for instance, a low IACC value improves the feeling of spaciousness and source width. Various stereo techniques based on decorrelation have also been used to create wider stereo images [8].

Other authors [9][10] have studied the apparent width of noise presented on two speakers and headphones. They found that the amount of correlation between the two presented channels had a dramatic impact on the perception of wideness. When uncorrelated noise signals are presented on two speakers, the noise seems to fill the complete space between the speakers while correlated signals produce a narrow sound source placed in between the speakers [9].

In all the studies mentioned above, wideness was always considered to be a one-dimensional attribute of sound sources. However, wideness can also be thought as having a two or three-dimensional aspect, forming a shape (Figure 1). Only one previous study [11] investigating sound shape perception is known to the author. In that work, the experiments were done binaurally on headphones and not using decorrelation; the results were mostly inconclusive.

The experiments presented in this paper are intended to assess the extent of source shape identification by subjects and also consider the general issues of source wideness in sound scenes. We first concentrate on sound shape creation.

62.1. Method for sound source shape creation

To create different sound source shapes we used a technique inspired by [9] and [12]. In the experiments, the sound source shapes were created using several point sources emitting uncorrelated signals. The reason for using decorrelated signals is that if identical or highly correlated signals were used, the different source signals would be summed and the listener's binaural system would perceive only a narrow sound image the position of which would depend on the placement and intensity of each sound source; this effect is widely used in panning techniques.

To create decorrelated point sources for a noise signal, statistically independent noise sequences were fed to each of the different sources. When the initial signal was a monoraul sound recording, a decorrelation filterbank was used.; this allowed the creation of several signals that were statistically uncorrelated but perceptually identical. The decorrelation filterbank can be implemented by several FIR or IIR all-pass filters that have different and randomized phase responses [9].

The decorrelation filterbank can also be made time-varying; this is known as dynamic decorrelation. Dynamic decorrelation is intended to emulate the micro-variations caused by moving air, temperature changes etc. and provides more natural sounding broad sound sources [9].

In the experiments, we used a filterbank of non time-varying all-pass FIR filters. The FIR filters had a 256-tap length and the filter coefficients were obtained using the frequency response sampling method. The magnitude response of the filters was 0 dB across all frequencies (all-pass) but their phase responses were all randomized and different.

This configuration allowed us to obtain several output signals that had correlation coefficients very close to zero between any two signals.

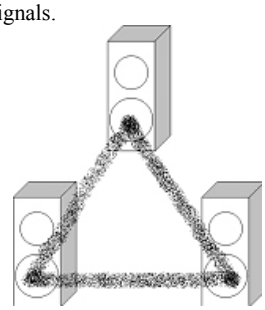

Figure 1. *Forming of a sound shape using real sound sources emitting uncorrelated signals.*

3. EXPERIMENTS

The experiments were carried out at the University of Wollongong, Australia, Thomson Corporate Research labs, Germany, and the Electrical and Telecommunication Research Institute (ETRI), Korea. The same apparatus and procedures were used in the three locations [13].

3.1. Real sound sources

In this experiment, subjects had to identify which shape was being played from six possible shapes. Eighteen test sequences were created which used the six shapes in a random order; each shape appearing exactly three times per experiment. No feedback was given to the subjects but, before starting the experiment, some initial training with the six different shapes was given.

Using the shape forming method previously described and switching on and off selected speakers of a specially arranged vertical 7-speaker array (Figure 2), different sound shapes were created. The different shapes were normalized to the same loudness. They are illustrated in Figure 3. The subject heads were at the same level as speaker 4 (looking straight at it or away from it depending on frontal or back experiment).

The experiment was repeated for four types of signals: white noise, 1 kHz low pass noise, 3 kHz high pass noise and a blues guitar riff. A further repetition was performed so as to present the speaker array both in front and behind the subjects at a distance of approximately 1.5 m.

Between 19 and 26 subjects took part in the experiments. Some subjects were audio engineers and had trained ears; some were not.

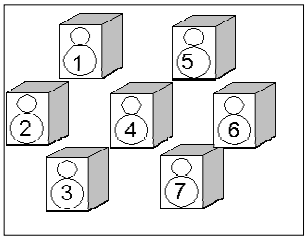

Figure 2. *Illustration of the 7-speaker array used in the real sound source experiment*

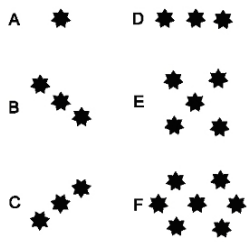

Figure 3. *The six different shapes used in the sequences*

Results

Table 1 shows the percentage of correct answers for each shape. Shape 'A' could be determined with almost 100 % confidence; this is likely to be because Shape 'A' was a point source and easy to discriminate against the other broad sound sources. Therefore, so as to not bias the results, only the correct choices between shapes 'B','C','D','E' and 'F' are shown. The statistical probability (p) of a selection was 20 % (5 shapes) and a double-sided significance test was performed to determine if the observed experiment results differed significantly from this statistical chance. The acceptance region was computed as:

$$ar = p \pm \frac{1.96 \cdot \sqrt{p(1-p)}}{\sqrt{n}} \qquad (1)$$

This is indicated in the column acceptance region (ar), which has to be interpreted in the following manner: If the average value is in the interval of the acceptance region, the statistical chance cannot be rejected. A level of significance of 95% was used. In Table 1, N is the total number of sequences for each

signal type while B and F indicate presentation of the shapes from the front and back, respectively.

		ETRI ***[%]***	**Th.** ***[%]***	**UoW** ***[%]***	***n***	***ar = 20±x [%]***	**Ave** ***[%]***
B	white noise	16.7	33.3	21.0	*390*	*20±4.0*	**23.6**
B	low p. noise	16.0	23.7	12.4	*390*	*20±4.0*	**17.7**
B	high p. noise	20.0	30.4	17.1	*390*	*20±4.0*	**22.8**
B	blues guitar	12.7	11.1	21.9	*390*	*20±4.0*	**14.6**
F	white noise	30.0	56.3	N/A	*285*	*20±4.6*	**42.5**
F	low p. noise	16.0	35.6	N/A	*285*	*20±4.6*	**23.5**
F	high p. noise	20.0	51.9	N/A	*285*	*20±4.6*	**41.4**
F	blues guitar	12.7	N/A	N/A	*150*	*20±6.4*	**17.3**

Table 1. *Percentage of correct choices between shapes 'B','C','D','E' and 'F' for the shape experiment*

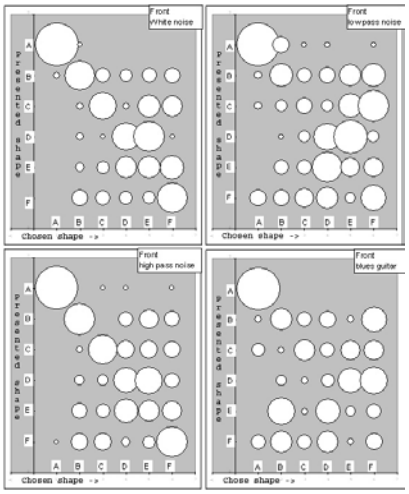

Figure 4. *Confusion matrices for white noise, low pass noise, high pass noise and blues guitar samples*

An analysis of the results in Table 1 thus indicates that satisfactory results were obtained only for white noise and high pass noise signals presented in front of subjects. It is well known that sound localization from behind is less accurate than from the front; therefore it seems that shape identification is correlated to localization accuracy as the results with the sound sources behind the listener are poor.

Another useful representation of the results is the so-called confusion matrix. In this representation, the shapes played to the subjects are placed on the vertical axis and the chosen shapes on the horizontal axis. Therefore, if identification of the various sound shapes is correct, we can expect to have the highest values on the diagonal of the matrix. This representation also shows the shapes that are the most often confused.

Figure 4 shows the confusion matrices for the shapes presented in the front for the four types of signal used.

3.2. Virtual sound sources

In contrast with the real sound source experiment, in the second experiment virtual sound sources were used to create the sound source shapes. The reason for performing this experiment was that practical speaker configurations cannot use the real sound source speaker array described previously. The use of virtual sound sources on headphones or speakers thus reflects a more realistic utilization scenario.

We used a first order Ambisonics [2] panning scheme on a cubic speaker array for this experiment. The test procedure was substantially the same as in the real sound source experiment. However, subjects had to choose between only five shapes present in ten sequences and some sequences used non-decorrelated point sources in order to prove that decorrelation helps to discriminate the shapes. We only used the white noise signal to maximize the chance of clear results. A total of sixteen subjects participated in the experiment.

Results

For decorrelated white noise, shapes were, on average, correctly selected 31.9% of the time (41.1% at Thomson, 20% at University of Wollongong). The statistical chance was 20% (5 shapes). Meanwhile, for correlated white noise, shapes were correctly selected 29.4% of the time (36.7 % at Thomson, 20% at the University of Wollongong).

Both sets of results were better than the statistical chance, but less than 50% were correct. It also seems that decorrelation helps in determining the shapes. The results are summarized in Table 2.

decorr.	**Th. [%]**	**UoW [%]**	***n***	***ar [%]***	**Ave [%]**
yes	41.1	20.0	*160*	*20 ±6.2*	**31.9**
no	36.7	20.0	*160*	*20 ±6.2*	**29.4**

Table 2. *Percentage of correct choices between shapes*

The decorrelated Thomson result is comparable with the result of the preceding real sound source test for white noise played in the front of the listener, which was 42.6%. This indicates that shape identification is very similar for both real and virtual sound sources. However, the result of the University of Wollongong was inside the significance interval and therefore we are not capable of drawing a conclusion.

3.2.1. Vertical and horizontal wideness

This experiment aimed at studying the perception of relative source wideness. For each of the test sequence, subjects were presented a narrow and a wider sound source. They were asked to specify which sound source was wider using a marking scheme from one to five (1: means the first source sounds much wider than the second; 3: the two sources have the same wideness and 5: the second source sounds much wider than the first). The angular extent of the sound sources ranged from 0 to 90 degrees. The experiment was performed once for horizontally extended sound sources and once for vertically extended sound sources. The results indicated that listeners

could compare different source wideness in the horizontal and vertical plane with a great precision even for small differences in wideness. Another experiment studied the perception of absolute apparent source width but space prevents to show these results.

3.3. Sound scenes

In the Sound scene experiment, subjects were asked to perform A-B comparisons between sound scenes that used broad sound sources and sound scenes that used solely point sources. The comparison criterion instructed to the subjects was *naturalness*. The sound samples used in the scenes referred to naturally large auditory events or objects (crowd, thunder, truck, beach, city and water).

The speaker system was the same as in the real sound source experiment and the order of playback of the broad and narrow scenes was randomised. The decorrelation technique was the same as in the shape experiment.

Results

The sound scenes that used broad sound sources were preferred 70.4% of the time; this significant result indicates that sound scenes using broad sound sources are perceived as being more natural sounding for representing large objects. This highlights the need for the implementation of wideness in virtual acoustic displays and MPEG-4 AudioBIFS. We are planning to carry out the opposite experiment as to verify if scenes using only point sources are perceived as being more natural sounding for familiarly narrow sounds (flying insect, a person's voice etc.).

3.4. Further experiments

We are currently carrying out further experiments into sound source wideness perception; in particular, these are intended to investigate other aspects of sound source wideness such as point source density, amount of correlation and dynamic decorrelation...Experiments aimed at studying the effect of dynamic decorrelation are considering, for example, the impact of the rate of change of the decorrelation on the perception of the sound source. We are also interested in understanding the effects of different correlation coefficients between point sources. Finally, point sources can also have movement - for example, the leaves and branches of a tree in wind produce constantly moving point sources; it is interesting to investigate what effects can be achieved by incorporating such features into reproduction models.

4. CONCLUSIONS

This paper has presented initial experiments into the perception and identification of sound source wideness and shape. Results showed that source horizontal and vertical wideness were clearly perceived by listeners and that sources with different wideness values could easily be discriminated. Further, sound scenes that used wide sound sources to represent large auditory objects were perceived as being more natural 70.4% of the time over scenes that used point only sources.

Identification of the source shape was, however, less reliable. In some cases, correct choice percentages were significantly above the statistical chance (for white noise and high pass noise presented from the front). This indicates that width and shape identification are highly dependent on the nature of the emitted signal.

With regard to MPEG-4 AudioBIFS, it seems that implementing a source shape description is overkill. We are however continuing investigation into the full set of parameters (e.g. point source density, dynamic decorrelation etc..) that are required to describe sound source wideness thoroughly in a virtual acoustic scene context. We thus expect AudioBIFS to include wideness parameters in the future.

Overall, shape and wideness of sound sources can be used to convey extra dimensions of information in auditory displays as well as improving realism of virtual sound scenes. We strongly believe that both shape and wideness merit further, and more detailed investigation.

5. ACKNOWLEDGEMENTS

The authors would like to thank the MPEG Audio subgroup and particularly Jens Spille of Thomson Central Research Labs, and Jeongil Seo of ETRI for their collaboration in these experiments.

6. REFERENCES

[1] ISO/IEC 14496-1 Information technology – Coding of audio-visual objects, Part 1: System

[2] Malham, D. G., Myatt A., "3-D Sound Spatialization using Ambisonic Techniques", Computer Music Journal, vol. 19(4), pp 58-10, 1995

[3] Boring, E.G., "Auditory theory with special reference to intensity, volume, and localization", American Journal of Psychology, vol. 37(2), pp 157-188, 1926

[4] Perrot, D., Buell, T., "Judgments of sound volume: Effects of signal duration, level, and interaural characteristics on the perceived extensity of broadband noise", J. Acoustical society of America, vol. 72(5), pp 1413-7, Dec. 1981

[5] Perrot, D., Musicant, A., Schwethelm, B., "The expanding image effect: The concept of tonal volume revisited", J. Auditory Research, vol. 20, pp 43-55, 1980

[6] Beranek, L., "Concert and opera halls: How they sound", J. Acoustical society of America, New York, 1996

[7] Blauert, J., "Spatial hearing", MIT press, 1998

[8] Gerzon, M., "Signal processing for simulating realistic stereo images", 93rd AES convention, New York, 1-4 October, Preprint 3424, 1992

[9] Kendall, G. S., "The Decorrelation of Audio Signals and Its Impact on Spatial Imagery", Computer Music Journal, vol. 19(4), pp 71-87, 1995

[10] Kurozumi, K., Ohgushi, K., "The relationship between the cross-correlation coefficient of two-channel acoustic signals and sound image quality", J. Acoustical society of America, vol. 74(6), pp 1726-1733, Dec. 1983.

[11] Hollander, J. S., "An exploration of virtual auditory shape perception", Masters Thesis, University of Washington, 1995

[12] Sensaura white papers, www.sensaura.com.

[13] Potard, G., Spille, J., "Study of Sound Source Shape and Wideness in Virtual and Real Auditory Displays", in proceedings of the AES 114th convention, Amsterdam, March 2003